BOUNDLESSINFORMANT – Frequently Asked Questions
09-06-2012

(U/FOUO) Questions

1) What is ***BOUNDLESSINFORMANT***? What is its purpose?
2) Who are the intended users of the tool?
3) What are the different views?
4) Where do you get your data?
5) Do you have all the data? What data is missing?
6) Why are you showing metadata record counts versus content?
7) Do you distinguish between sustained collect and survey collect?
8) What is the technical architecture for the tool?
9) What are some upcoming features/enhancements?
10) How are new features or views requested and prioritized?
11) Why are record counts different from other tools like ASDF and What's On Cover?
12) Why is the tool NOFORN? Is there a releasable version?
13) How do you compile your record counts for each country?

Note: This document is a work-in-progress and will be updated frequently as additional questions and guidance are provided.

1) (U) What is *BOUNDLESSINFORMANT*? What is its purpose?

(U//FOUO) BOUNDLESSINFORMANT is a GAO prototype tool for a self-documenting SIGINT system. The purpose of the tool is to fundamentally shift the manner in which GAO describes its collection posture. ***BOUNDLESSINFORMANT*** provides the ability to dynamically describe GAO's collection capabilities (through metadata record counts) with no human intervention and graphically display the information in a map view, bar chart, or simple table. Prior to ***BOUNDLESSINFORMANT***, the method for understanding the collection capabilities of GAO's assets involved ad hoc surveying of repositories, sites, developers, and/or programs and offices. By extracting information from every DNI and DNR metadata record, the tool is able to create a near real-time snapshot of GAO's collection capability at any given moment. The tool allows users to select a country on a map and view the metadata volume and select details about the collection against that country. The tool also allows users to view high level metrics by organization and then drill down to a more actionable level - down to the program and cover term.

Sample Use Cases

- (U//FOUO) How many records are collected for an organizational unit (e.g. FORNSAT)?
- (U//FOUO) How many records (and what type) are collected against a particular country?
- (U//FOUO) Are there any visible trends for the collection?
- (U//FOUO) What assets collect against a specific country? What type of collection?
- (U//FOUO) What is the field of view for a specific site? What countriees does it collect against? What type of collection?

2) (U) Who are the intended users of the tool?

- (U//FOUO) Mission and collection managers seeking to understand output characteristics of a site based on what is being ingested into downstream repositories.
- (U//FOUO) Strategic Managers seeking to understand top level metrics at the organization/office level or seeking to answer data calls on NSA collection capability.

- (U//FOUO) Analysts looking for additional sites to task for coverage of a particular technology within a specific country.

3) What are the different views?

(U//FOUO) Map View – The Map View is designed to allow users to view overall DNI, DNR, or aggregated collection posture of the agency or a site. Clicking on a country will show the collection posture (record counts, type of collection, and contributing SIGADs or sites) against that particular country in addition to providing a graphical display of record count trends. In order to bin the records into a country, a normalized phone number (DNR) or an administrative region atom (DNI) must be populated within the record. Clicking on a site (within the Site Specific view) will show the viewshed for that site – what countries the site collects against.

(U//FOUO) Org View – The Organization View is designed to allow users to view the metadata record counts by organizational structure (i.e. GAO – SSO – RAM-A – SPINNERET) all the way down to the cover term. Since it's not necessary to have a normalized number or administrative region populated, the numbers in the Org View will be higher than the numbers in the Map View.

(U//FOUO) Similarity View – The Similarity View is currently a placeholder view for an upcoming feature that will graphically display sites that are similar in nature. This can be used to identify areas for a de-duplication effort or to inform analysts of additional SIGADs to task for queries (similar to Amazon's "if you like this item, you'll also like these" feature).

4) (U) Where do you get your data?

(U//FOUO) BOUNDLESSINFORMANT extracts metadata records from GM-PLACE post-FALLOUT (DNI ingest processor) and post-TUSKATTIRE (DNR ingest processor). The records are enriched with organization information (e.g. SSO, FORNSAT) and cover term. Every valid DNI and DNR metadata record is aggregated to provide a count at the appropriate level. See the different views question above for additional information.

5) (U) Do you have all the data? What data is missing?

- (U//FOUO) The tool resides on GM-PLACE which is only accredited up to TS//SI//NOFORN. Therefore, the tool does not contain ECI or FISA data.
- (U//FOUO) The Map View only shows counts for records with a valid normalized number (DNR) or administrative region atom (DNI).
- (U//FOUO) Only metadata records that are sent back to NSA-W through FASCIA or FALLOUT are counted. Therefore, programs with a distributed data distribution system (e.g. MUSCULAR and Terrestrial RF) are not currently counted.
- (U//FOUO) Only SIGINT records are currently counted. There are no ELINT or other "INT" records included.

6) (U) Why are you showing metadata record counts versus content?

(U//FOUO)

7) (U) Do you distinguish between sustained collect and survey collect?

(U//FOUO) The tool currently makes no distinction between sustained collect and survey collect. This feature is on the roadmap.

8) What is the technical architecture for the tool?

- Click here for a graphical view of the tool's architecture
- (U//FOUO) DNI metadata (ASDF), DNR metadata (FASCIA) delivered to Hadoop Distributed File System (HDFS) on GM-PLACE
- (U//FOUO) Use Java MapReduce job to transform/filter and enrich FASCIA/ASDF data with business logic to assign organization rules to data
- (U//FOUO) Bulk import of DNI/DNR data (serialized Google Protobuf objects) into Cloudbase (enabled by custom aggregators)
- (U//FOUO) Use Java web app (hosted via Tomcat) on MachineShop (formerly TurkeyTower) to query Cloudbase
- (U//FOUO) GUI triggers queries to CloudBase – GXT (ExtGWT)

9) What are some upcoming features/enhancements?

- (U//FOUO) Add technology type (e.g. JUGGERNAUT, LOPER) to provide additional granularity in the numbers
- (U//FOUO) Add additional details to the Differential view
- (U//FOUO) Refine the Site Specific view
- (U//FOUO) Include CASN information
- (U//FOUO) Add ability to export data behind any view (pddg,sigad,sysid,casn,tech,count)
- (U//FOUO) Add in selected (vs. unselected) data indicators
- (U//FOUO) Include filter for sustained versus survey collection

10) How are new features or views requested and prioritized?

(U//FOUO) The team uses Flawmill to accept user requests for additional functionality or enhancements. Users are also allowed to vote on which functionality or enhancements are most important to them (as well as add comments). The ***BOUNDLESSINFORMANT*** team will periodically review all requests and triage according to level of effort (Easy, Medium, Hard) and mission impact (High, Medium, Low). The team will review the queue with the project champion and government steering committee to be added onto the ***BOUNDLESSINFORMANT*** roadmap.

11) Why are record counts different from other tools like ASDF and What's On Cover?

(U//FOUO) There are a number of reasons why record counts may vary. The purpose of the tool is to provide